Rethinking the Intelligence Quotient: A critical analysis of the foundations, limitations and sociopolitical implications of intelligence tests

Ernesto Mascarell Perez
Independent Researcher
mascarell1960@hotmail.com

Abstract

More than a century after the creation of the first practical intelligence test, there remains remarkable scientific disagreement on how human intelligence should be described, whether IQ tests truly measure it and, if not, what they actually evaluate. This article offers a critical and systematic review of the epistemological, methodological and ethical foundations of intelligence psychometrics. Four central problem areas are examined: (1) the nature and stability of the construct "intelligence" and the *g* factor; (2) the cultural and environmental bias inherent in measurement instruments; (3) the methodological deficiencies in heritability research designs; and (4) the social and ethical consequences of applying these tests in educational, occupational and clinical contexts. It is argued that the claim of IQ tests to measure a general, innate and hierarchical cognitive capacity lacks solid empirical foundation and constitutes, to a large extent, a statistical and cultural artifact. The article concludes with recommendations for future research and professional practice, advocating for alternative models of assessing human potential.

Keywords: Intelligence Quotient, intelligence tests, *g* factor, cultural bias, heritability, psychometrics, biological determinism, validity.

1. Introduction

The measurement of human intelligence constitutes one of the most ambitious and controversial projects of scientific psychology. Since Alfred Binet and Théodore Simon developed the first practical scale in 1905, intelligence tests have spread globally, being used in educational contexts for identifying "gifted" and "intellectually disabled" individuals, in the workplace for personnel selection, and in clinical settings for diagnosing various conditions. This ubiquity contrasts, however, with the persistence of fundamental debates about the

nature of the construct they purport to measure and the validity of their applications.

Ken Richardson (2002) clearly stated the central paradox: "As we approach the centenary of the first practical intelligence test, there is still little scientific agreement on how human intelligence should be described, whether IQ tests really measure it and, if not, what they really measure." This uncertainty is not mere academic technicality; it has profound implications for social mobility, educational equity and distributive justice.

The aim of this article is twofold. On the one hand, it proposes to offer a critical synthesis of the empirical evidence that questions the fundamental assumptions of intelligence psychometrics. On the other, it seeks to examine the practical consequences of applying these tests, arguing that their uncritical use tends to reinforce structural inequalities under a veneer of scientific neutrality.

2. The construct problem: What is intelligence?

2.1 The reification of the *g* factor

The concept of general intelligence, or *g* factor, introduced by Charles Spearman at the beginning of the 20th century, remains the theoretical pillar of most contemporary IQ tests. Spearman observed that scores on different cognitive tests tended to correlate positively—the so-called "positive manifold"—and inferred the existence of a common factor underlying all of them.

Paleontologist Stephen Jay Gould, in his work The Mismeasure of Man (1981), subjected this reasoning to fundamental criticism. Gould argued that the *g* factor is a product of what he termed the "fallacy of reification": the tendency to convert abstract concepts into real entities. The correlation between tests does not demonstrate the existence of a general mental capacity; it can equally be explained by methodological, experiential or sociocognitive factors that affect performance on any test task.

2.2 The problems of CHC Theory

The Cattell-Horn-Carroll (CHC) model represents the most sophisticated attempt to systematize the structure of cognitive abilities. This three-stratum hierarchical model—with a *g* factor at the apex, multiple broad abilities (such as crystallized intelligence Gc and fluid intelligence Gf) at an intermediate level,

and specific abilities at the base—has been widely adopted as a framework for the development and interpretation of intelligence tests.

However, even its defenders recognize substantial limitations. John D. Wasserman (2019) has pointed out that the CHC model suffers from "weak explanatory power, its inability to make testable predictions, its entanglement with a single measurement instrument (the Woodcock-Johnson), inconsistent independent replication of its factor structure and, most importantly, its unproven evidentiary value for the identification and intervention with students with specific learning disorders." Wasserman adds that these limitations are inherent to any psychological theory developed or tested primarily from factor analyses.

2.3 Theoretical alternatives and their empirical status

In contrast to the unitary model, multidimensional proposals have emerged. Howard Gardner's theory of multiple intelligences (1983), which postulates the existence of at least eight relatively independent cognitive domains, has enjoyed great popularity in educational circles. However, its scientific status is problematic. A recent review published in Frontiers in Psychology (2023) concludes that "many researchers have claimed that Gardner's theory of multiple intelligences is a neuromyth because they have not seen evidence supporting his proposal of brain-based independent intelligences." Factorial studies have not demonstrated that intelligences are independent, and research on the effects of MI-based teaching has not explored alternative causes for positive effects nor been conducted using standard scientific methods.

For his part, Sternberg's theory of successful intelligence (1999) proposes a triarchic model (analytical, creative and practical). Sternberg himself has criticized both *g* factor theorists—for using restricted ranges of participants, materials or situational contexts—and defenders of new multiple intelligences—for subjecting their theories to little or no empirical testing—advocating for a middle ground that recognizes the multifaceted nature of intelligence.

This fragmented landscape suggests that, far from there being a consensus on the nature of intelligence, multiple conceptual frameworks coexist with different degrees of empirical support, none of which can be considered definitive.

3. Cultural and environmental bias in measurement

3.1 The Flynn Effect and historical plasticity

One of the most disturbing findings for the conception of intelligence as a fixed and innate trait is the Flynn Effect: the sustained increase in IQ scores throughout the 20th century. In 1981, psychologist James Flynn observed that IQ scores had been steadily rising for almost a century, with a surprising difference of 18 points over two generations. After careful analysis, Flynn concluded that the cause was cultural: society had become more intelligent in terms of handling larger and more abstract ideas, and had made people more intelligent.

This finding directly undermines the idea of intelligence as a fixed and innate entity. As David Shenk points out, "this observation, combined with solid evidence that IQ scores are also not fixed within an individual, clearly dispels the idea that intelligence is an innate and fixed entity." The largest increases in scores over time tend to be associated with a reduction in culturally oriented content in IQ tests, indicating that even tests designed to be "culture-free" are not completely so.

3.2 Evidence of bias in visuospatial reasoning tests

Even tests presented as "culture-fair," such as Raven's Progressive Matrices or Cattell's culture-fair test, do not escape cultural bias. A 2022 integrative review concluded that "several deeply ingrained cultural assumptions appear in all visuospatial reasoning tests, to the point that it invalidates the view of such tests as intrinsically culture-fair and makes it impossible to draw clear conclusions from mean score differences between ethnic groups."

A systematic review of 225 studies published in 2025 identified ten factors contributing to group differences on cognitive ability tests, including: biological and environmental factors, test anxiety, attitudes, experience and stereotype threat. The results showed that, while biological and environmental factors are related to group differences, so are several factors relevant to test design and administration, "most notably stereotype threat, differences in experience and domain exposure, and self-confidence."

3.3 The radical critique: IQ tests as proxies for social class

Ken Richardson has taken the critique a step further, arguing that "the entire population variance in IQ scores can be described in terms of a nexus of sociocognitive-affective factors that differentially prepare individuals for the cognitive, affective and performance demands of the test—in effect, that the test is a measure of social class origin, and not of capacity for complex cognition

as such." Richardson demonstrates how these factors can explain the correlational evidence generally considered to validate IQ tests, including associations with educational attainment, occupational performance and elementary cognitive tasks.

This critique does not imply denying individual differences in cognitive abilities, but rather questioning that IQ tests validly measure them and that observed differences reflect a stable underlying capacity rather than accumulated inequalities in opportunities, resources and experiences.

4. The inheritance of intelligence: Methodological review

4.1 The heritability debate

One of the most recurrent arguments in favor of the validity of IQ tests is their supposed high heritability, often estimated between 50% and 80%. A recent study published in Scientific Reports (2026) reported an IQ heritability of approximately 75%, with genetic factors explaining most of the association between IQ and socioeconomic status (69-98%).

However, these estimates must be interpreted with caution. Heritability is a population parameter that varies according to context and does not imply that individual differences are genetically determined in a fixed manner. In fact, some adoption studies suggest that shared environment could explain up to 50% of the variance in IQ when corrected for range restriction.

4.2 Gene-Environment Interactions

Research on gene-environment interactions adds an additional layer of complexity. A meta-analysis by Tucker-Drob and colleagues (2015) found that the interaction between socioeconomic status and IQ heritability was not consistent across countries. While some US studies reported that heritability increased with higher SES levels, this pattern was "generally absent in studies outside the US (England, Europe, Australia)." The authors speculated that this difference might reflect "worse conditions for intellectual development of low SES individuals in the US than in other countries."

An analysis of data from the Texas and Colorado Adoption Projects found no substantial heritability x SES interaction, with effect size estimates negative or

close to zero. This suggests that SES may not fully capture the proximal environmental aspects that moderate childhood IQ.

4.3 The issue of range restriction

A crucial methodological problem in adoption studies is the range restriction of the family environment. Families that adopt children tend to be disproportionately middle-to-upper SES, meaning that the range of environments studied does not represent the total variability of the general population. A correction for this restriction suggests that "shared environment could explain up to 50% of the variance in IQ." This seriously questions the claim that genetic factors explain most of the variance in IQ.

5. Practical and ethical implications

5.1 Predictive validity in the workplace

The supposed predictive validity of IQ tests for job performance has been a key argument for their use in personnel selection. However, recent evidence questions this claim. An updated meta-analysis by Berry, Lievens, Zhang and Sackett (2024) revealed that "the criterion validity of GMA [general mental ability] tests has been considerably overestimated due to inappropriate range restriction corrections." Their results lead to the conclusion that "excluding GMA tests generally has little or no effect on validity, but substantially decreases adverse impact."

A critical review published in the Journal of Occupational and Organizational Psychology (2023) examined five critical observations regarding the use of cognitive ability tests in selection for graduate and professional occupations: (1) evidence of adverse impact on test outcomes; (2) the tendency to position cognitive ability tests early in high-volume selection processes; (3) recent evidence challenging the meta-analytic validity of cognitive ability tests; (4) weaknesses in historical primary validity studies; and (5) a conceptually flawed examination of differential validity evidence. The authors conclude that "the practice of applying cognitive ability tests in these contexts has strong potential to maintain and exacerbate social inequality in access to higher occupations and professions."

5.2 Clinical and educational use

IQ tests are used in many countries to determine eligibility for special education, access to health services and diagnosis of intellectual disability. These decisions have lifelong consequences for individuals. However, given the documented cultural and environmental bias in these instruments, there is a real risk that they mislabel children from disadvantaged backgrounds as "disabled," while children from privileged backgrounds are identified as "gifted"—not due to differences in innate potential, but due to differences in opportunities.

5.3 Alternatives: Dynamic assessment

In response to the limitations of static tests (such as traditional IQ tests), alternatives have emerged that focus on learning potential rather than past achievements. Dynamic assessment, associated with the work of Reuven Feuerstein and Robert Sternberg, emphasizes "learning potential rather than past learning achievements." This approach, which follows a test-train-retest format, measures the individual's ability to benefit from instruction and improve with help—an approach closer to what many understand by "potential" than a static measurement of current level.

6. Conclusion and recommendations

The evidence reviewed in this article suggests that IQ tests, far from being neutral instruments that measure a universal and fixed cognitive capacity, are culturally and historically situated artifacts whose scores reflect, to a non-negligible extent, inequalities in educational opportunities, sociocultural experiences and material resources. The *g* factor, rather than a discovery about the architecture of the human mind, appears to be a statistical artifact whose apparent stability hides a profound contextual dependency.

This does not imply denying the existence of individual differences in cognitive abilities. It implies, rather, questioning the claim that IQ tests validly and fairly measure them, and rejecting the inference that observed differences reflect a hierarchical innate general capacity.

Based on this review, the following recommendations are formulated:

1. For researchers: Abandon the search for a single universal *g* factor in favor of multifactorial and contextually situated models. Prioritize the study of cognitive plasticity and the conditions that facilitate it over the measurement of supposed fixed traits.
2. For professionals (educators, clinical psychologists, selectors): Exercise maximum caution in interpreting IQ scores, explicitly considering the sociocultural context of the assessed individual and the instrument's

limitations. Seriously consider replacing static IQ tests with dynamic assessment approaches that measure learning potential.
3. For policymakers: Refrain from using IQ tests as a basis for high-impact decisions (access to special education, job selection) without rigorous justification demonstrating their validity for the specific population and their superiority over less biased alternatives.
4. For professional training: Incorporate critical training on the history, foundations and limitations of intelligence tests in psychology, education and human resource management curricula, promoting a nuanced understanding that recognizes both their potential utilities and their serious limitations.

7. Transparency declaration

The author declares no conflicts of interest relevant to the content of this article. The research underlying this work has been based exclusively on published and publicly accessible sources, and has not received specific funding.

References

1. Flynn, J. R. (1987). Massive IQ gains in 14 nations: What IQ tests really measure. Psychological Bulletin, 101(2), 171-191.
2. Gardner, H. (1983). Frames of mind: The theory of multiple intelligences. Basic Books.
3. Gould, S. J. (1981). The mismeasure of man. W. W. Norton & Company.
4. Herrmann, E., et al. (2012). Are there geniuses among the apes? Philosophical Transactions of the Royal Society B, 367(1603), 2753-2761.
5. Hunt, E., & Carlson, J. (2007). Considerations relating to the study of group differences in intelligence. Perspectives on Psychological Science, 2(2), 194-213.
6. Jensen, A. R. (1980). Bias in mental testing. Free Press.
7. McGill, R. J., & Dombrowski, S. C. (2019). Critically reflecting on the origins, evolution, and impact of the Cattell-Horn-Carroll (CHC) model. Applied Measurement in Education, 32(3), 216-231.
8. Richardson, K. (2002). What IQ tests test. Theory & Psychology, 12(3), 283-314.
9. Richardson, K., & Norgate, S. (2015). Does IQ really predict job performance? Psychological Science, 26(2), 145-155.
10. Rosenthal, R., & Jacobson, L. (1968). Pygmalion in the classroom: Teacher expectation and pupils' intellectual development. Holt, Rinehart & Winston.

11. Sternberg, R. J. (1999). Successful intelligence: Finding a balance. Trends in Cognitive Sciences, 3(11), 436-442.
12. Tucker-Drob, E. M., & Bates, T. C. (2016). Large cross-national differences in gene × socioeconomic status interaction on intelligence. Psychological Science, 27(2), 138-149.
13. Wasserman, J. D. (2019). Deconstructing CHC. Applied Measurement in Education, 32(3), 232-248.
14. Wilson, C. J., et al. (2023). The cross-cultural generalizability of cognitive ability measures: A systematic literature review. Intelligence, 98, 101751.
15. Woods, S. A., & Patterson, F. (2023). A critical review of the use of cognitive ability testing for selection into graduate and higher professional occupations. Journal of Occupational and Organizational Psychology.
16. Cuppello, S., Zibarras, L. D., & Corr, P. J. (2025). Factors related to mean score group differences on cognitive ability tests: A systematic review. Trends in Psychology.